National Cancer Institute

caBIG™

cancer Biomedical Informatics Grid™

The caBIG™ Data Sharing and Security Framework

Webcast for NCI Cancer Centers
May 16, 2008

Today’s Presentation

- Welcome
- Overview
- Introduction to the DSSF Bundle
- caBIG™ Trust Fabric
- DSSF Decision Support Tools
- Other DSSF Resources
- DSSF Future State
- Q&A

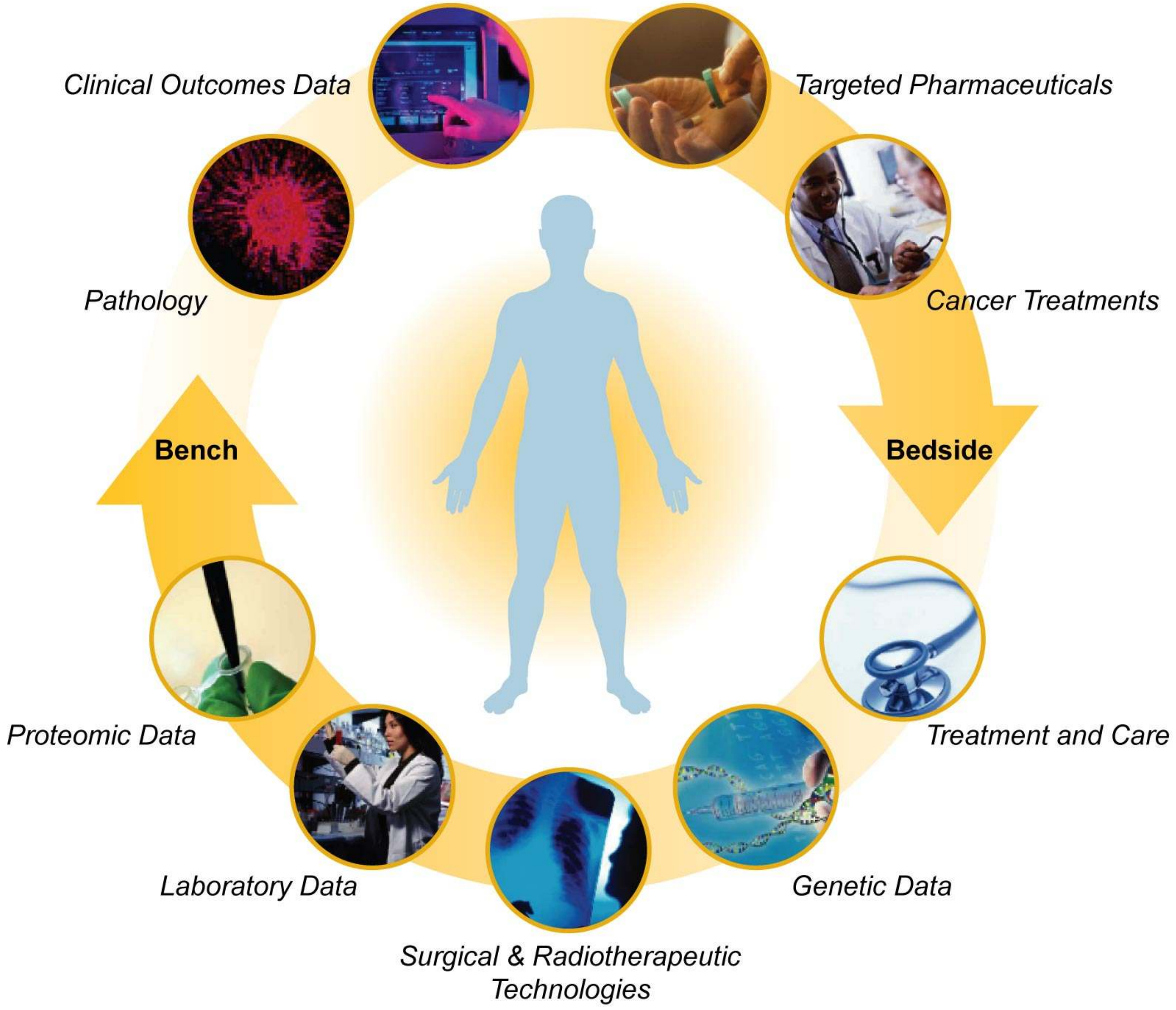
National Cancer Institute

caBIG™

cancer Biomedical Informatics Grid ™

Overview

Individualized, Targeted Cancer Care



The caBIG™ Initiative

caBIG™ Vision

A virtual network of interconnected data, individuals, and organizations that whose goal is to redefine how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.

caBIG™ Goals

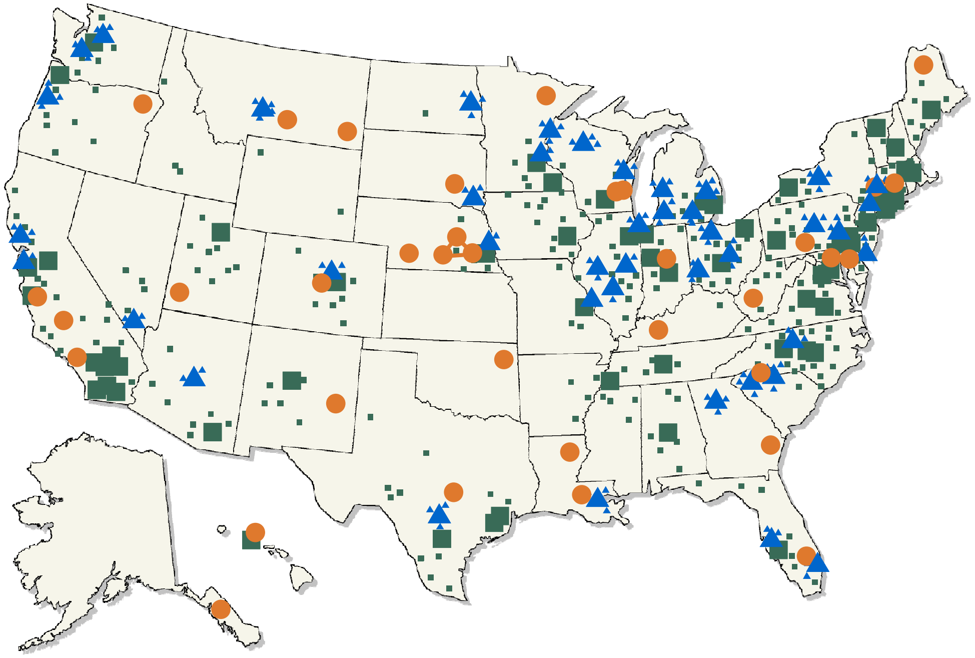
- **Connect** the cancer research community through a shareable, interoperable electronic infrastructure
- **Deploy and extend** standard rules and a common language to more easily share information
- **Build** or adapt tools for collecting, analyzing, integrating and disseminating information associated with cancer research and care

caBIG™ Progress to Date

- **caBIG™ provides:**
  - 40+ applications in discovery, clinical trials management, biospecimen management, etc.
  - Available through:
    - Clinical Trials Compatibility Framework
    - Life Sciences Distribution
    - Data Sharing and Security Framework
  - A connectivity infrastructure (caGrid)

- **caBIG™ adoption is unfolding in:**
  - 46 NCI-designated Cancer Centers
  - 16 NCI Community Cancer Centers
- caBIG™ being integrated into federal health architecture to connect **National Health Information Network**

NCI-Designated Cancer Centers and Community Cancer Centers



What is a Grid?

- **Grids have evolved from the concept of distributed computing to support science and engineering.**

- **Key Features and Benefits:**

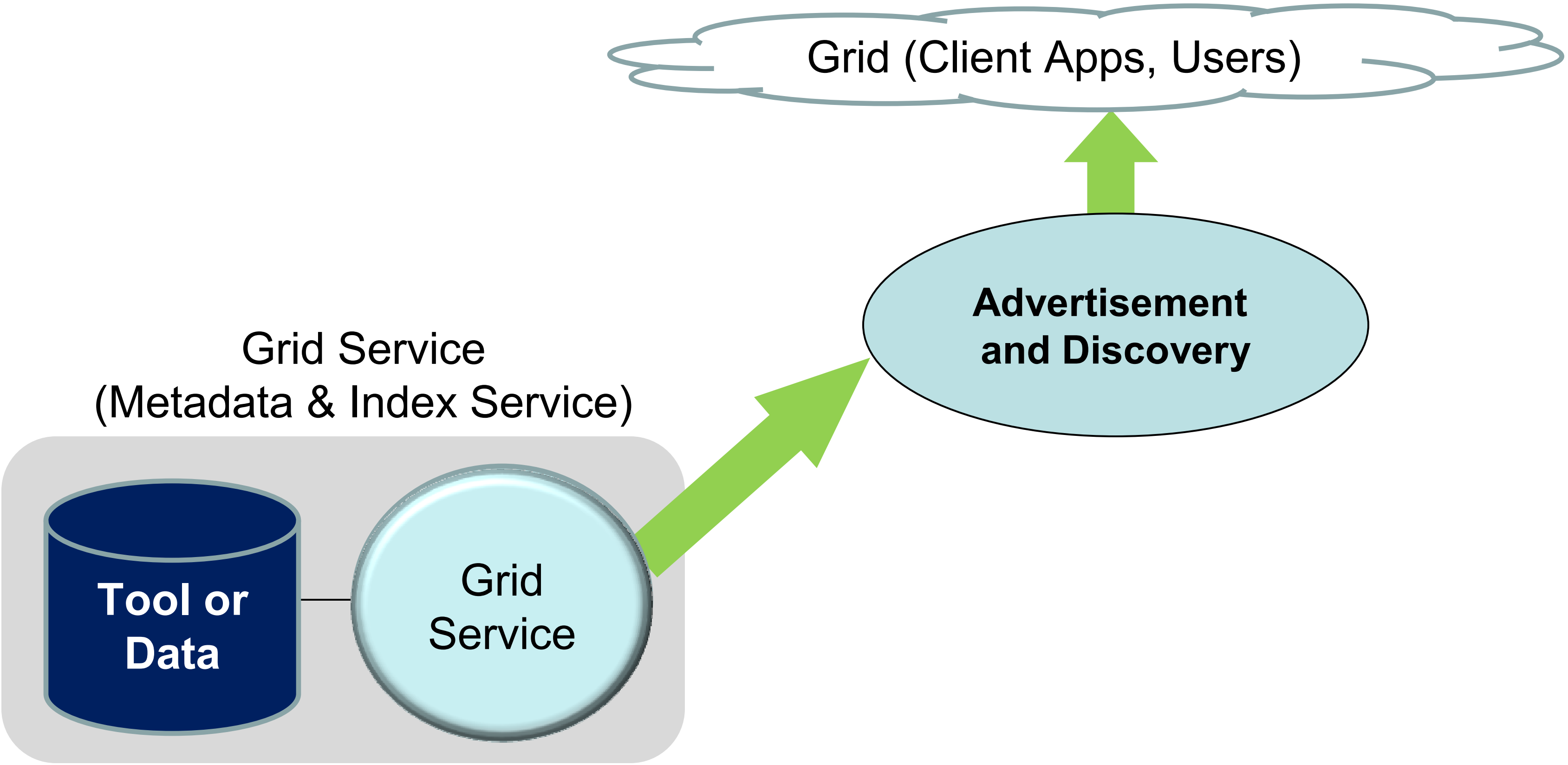
  - Sharing of resources (computational, storage, data, etc)
  - Secure Access (global authentication, local authorization, policies, trust, etc)
  - Open Standards
  - Virtualization

> *“The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations.”*
>
> *I. Foster, C. Kesselman, S. Tuecke. International J. Supercomputer Applications, 15(3), 2001.*

Source: caBIG Annual Meeting 2007: caGrid 1.0 Tutorial Overview

Grids Help Users Find Services & Data

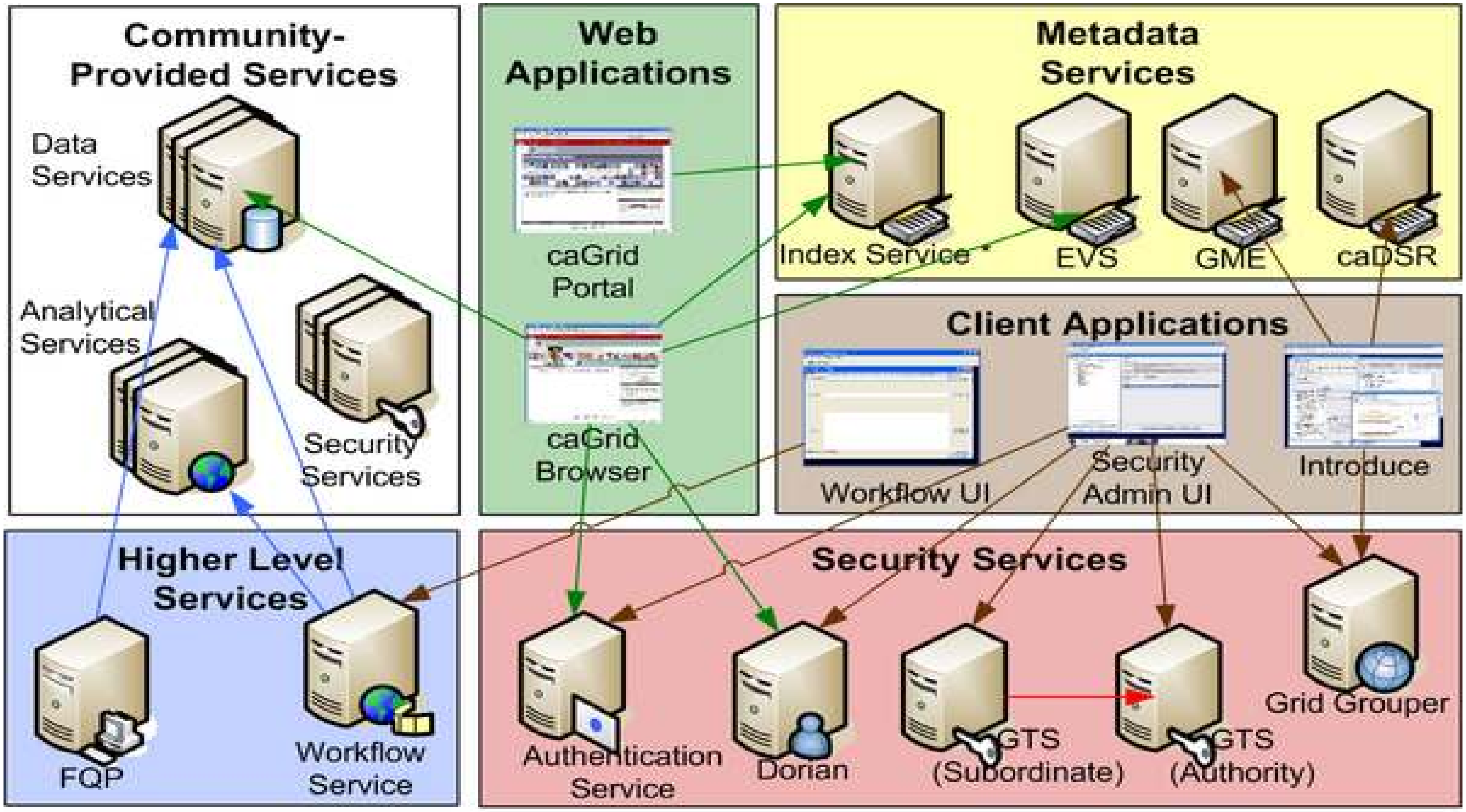
- **Metadata (information about the stored data) is deposited in a “Grid index service” that can be queried by grid users (Advertisement and Discovery).**


Grid (Client Apps, Users)
Advertisement and Discovery
Grid Service
(Metadata & Index Service)
Tool or Data
Grid Service

Source: Modified from caBIG Annual Meeting 2007: caGrid 1.0 Tutorial Overview

caGrid Infrastructure & Tooling: High Level View

- **Ultimately, the Grid consists of a collection of applications and services, connected to each other through a secure infrastructure.**



Source: www.cagrid.org

*All Services Register with the Index Service

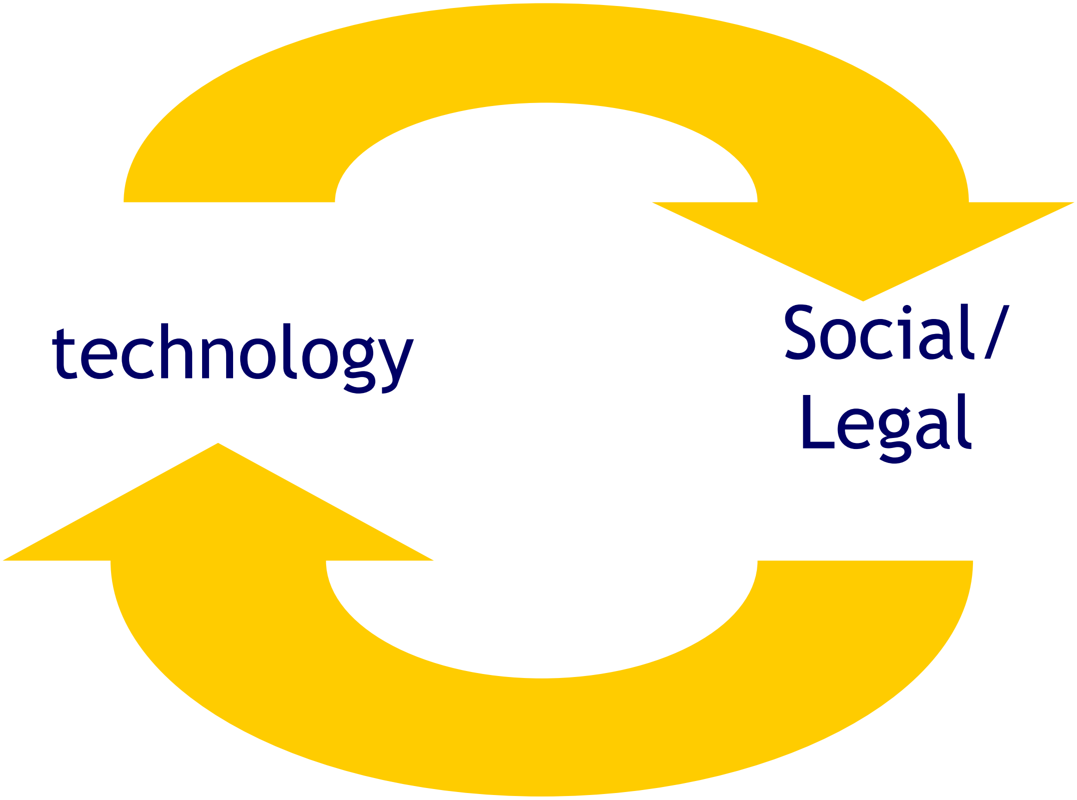
National Cancer Institute

caBIG™

cancer Biomedical Informatics Grid™

Introduction to the DSSF Bundle

Disambiguating social and technology issues



technology

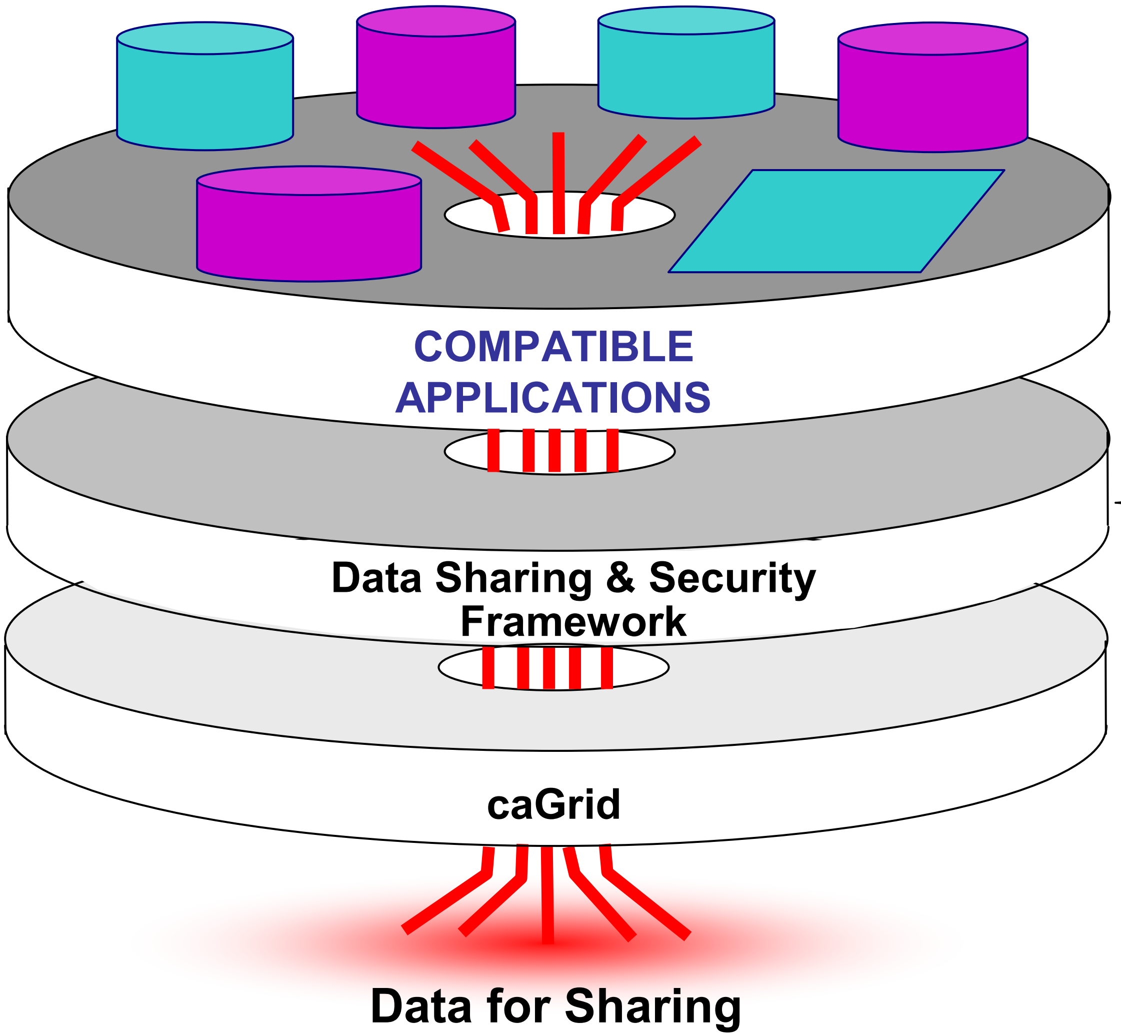
Social/
Legal

Removing Barriers to Data Sharing

- **Expectations for data sharing in caBIG™ need to be managed within a framework of established regulatory requirements, IP rights and existing incentive structures.**

- **Maximum utility of the caBIG™ infrastructure depends on addressing potential (or perceived) legal, regulatory and socio-cultural barriers.**

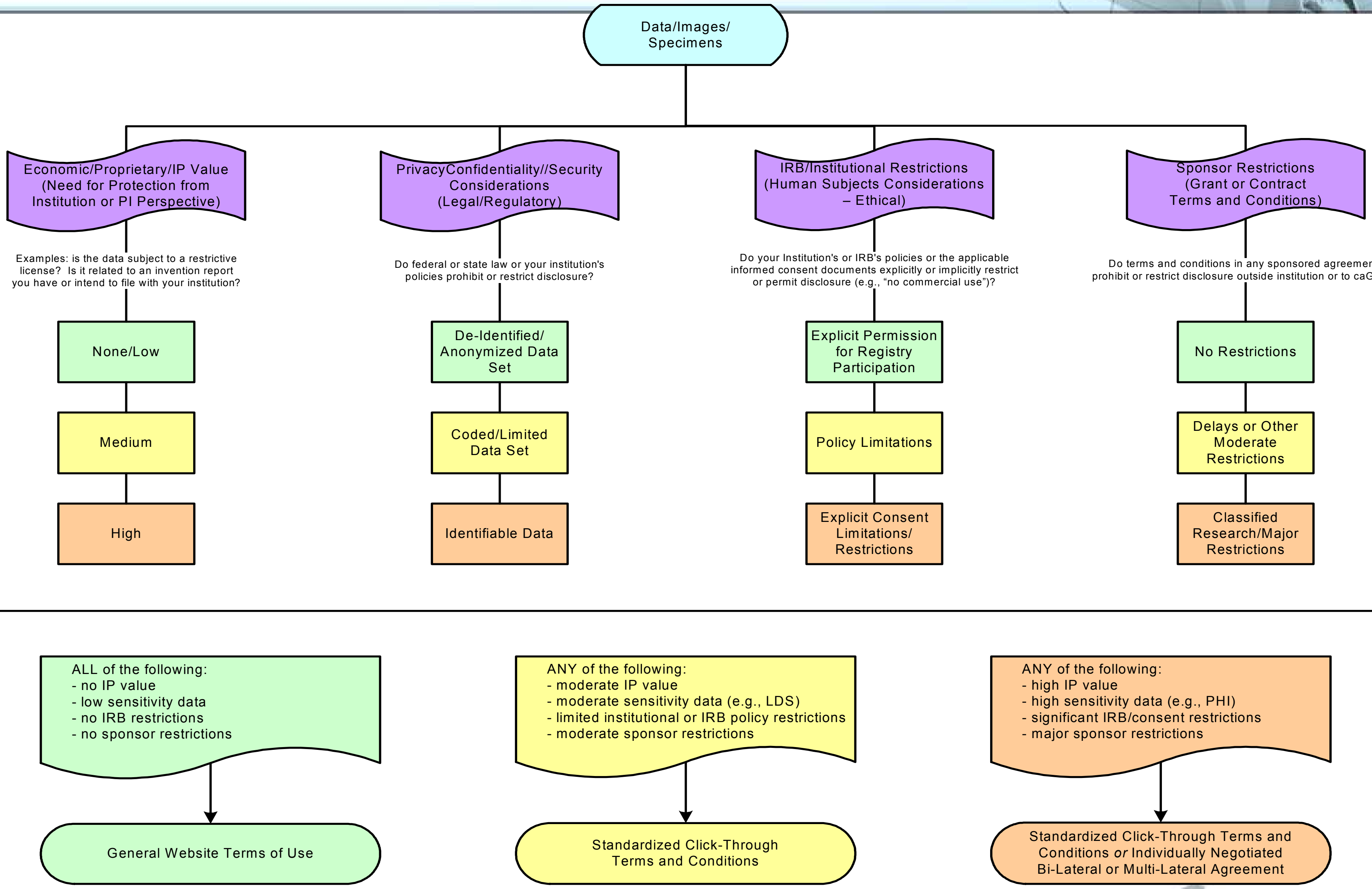
caBIG™ Data Sharing and Security Framework (DSSF)



When fully developed, the caBIG™ Data Sharing and Security Framework will include:

- caBIG™ Policies
- Processes and Best Practices
- Model Documents
- Trust Fabric

caBIG™ Data Sharing and Security Framework (DSSF)



Data/Images/ Specimens

Economic/Proprietary/IP Value (Need for Protection from Institution or PI Perspective)

Examples: is the data subject to a restrictive license? Is it related to an invention report you have or intend to file with your institution?

None/Low

Medium

High

PrivacyConfidentiality//Security Considerations (Legal/Regulatory)

Do federal or state law or your institution's policies prohibit or restrict disclosure?

De-Identified/ Anonymized Data Set

Coded/Limited Data Set

Identifiable Data

IRB/Institutional Restrictions (Human Subjects Considerations – Ethical)

Do your Institution's or IRB's policies or the applicable informed consent documents explicitly or implicitly restrict or permit disclosure (e.g., “no commercial use”)?

Explicit Permission for Registry Participation

Policy Limitations

Explicit Consent Limitations/ Restrictions

Sponsor Restrictions (Grant or Contract Terms and Conditions)

Do terms and conditions in any sponsored agreements prohibit or restrict disclosure outside institution or to caGRID?

No Restrictions

Delays or Other Moderate Restrictions

Classified Research/Major Restrictions

ALL of the following:
- no IP value
- low sensitivity data
- no IRB restrictions
- no sponsor restrictions

General Website Terms of Use

ANY of the following:
- moderate IP value
- moderate sensitivity data (e.g., LDS)
- limited institutional or IRB policy restrictions
- moderate sponsor restrictions

Standardized Click-Through Terms and Conditions

ANY of the following:
- high IP value
- high sensitivity data (e.g., PHI)
- significant IRB/consent restrictions
- major sponsor restrictions

Standardized Click-Through Terms and Conditions *or* Individually Negotiated Bi-Lateral or Multi-Lateral Agreement

Using the DSSF to Determine What Data Can Be Shared and How

- **Use the Framework's sensitivity analysis process to**
  - Determine which data can be shared –
  - Identify necessary access and data security controls (authentication, authorization)
  - ***The institution providing the data makes this determination.***

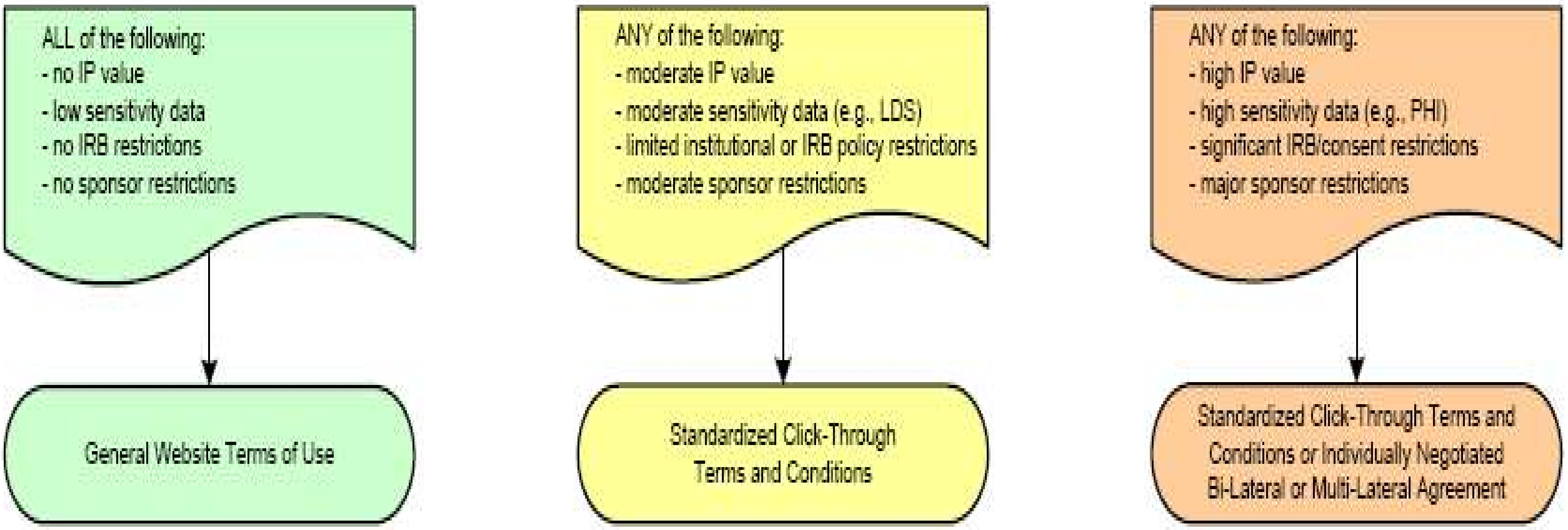
- **Audiences**
  - IRBs
  - Privacy Officials
  - Industry-Sponsored Projects/Grants & Contracts Administration/Tech Transfer Officers
  - Institutional Attorneys

Use the DSSF to Assess Sensitivity of Individual Datasets

- **Sensitivity Analysis Process**
  - Assess data sensitivity by reference to the Framework’s four principal elements:
    - **Economic/Proprietary Value (to Researcher/Institution)**
    - **Privacy Considerations**
    - **IRB/Ethical Restrictions**
    - **Sponsor Restrictions or Requirements**
  - Assign a **low**, **medium** or **high** sensitivity rating to the data
  - Review the outcome of the sensitivity analysis to determine the type of agreement suggested and the security/data access controls associated with the outcome - **Green**, **Yellow** or **Orange** levels of data

Use the DSSF to Select Type of Agreement and Access Controls

After conducting the sensitivity assessment, the providing researcher/institution can then select an appropriate data sharing mechanism.



ALL of the following:
- no IP value
- low sensitivity data
- no IRB restrictions
- no sponsor restrictions

General Website Terms of Use

ANY of the following:
- moderate IP value
- moderate sensitivity data (e.g., LDS)
- limited institutional or IRB policy restrictions
- moderate sponsor restrictions

Standardized Click-Through Terms and Conditions

ANY of the following:
- high IP value
- high sensitivity data (e.g., PHI)
- significant IRB/consent restrictions
- major sponsor restrictions

Standardized Click-Through Terms and Conditions or Individually Negotiated Bi-Lateral or Multi-Lateral Agreement

National Cancer Institute

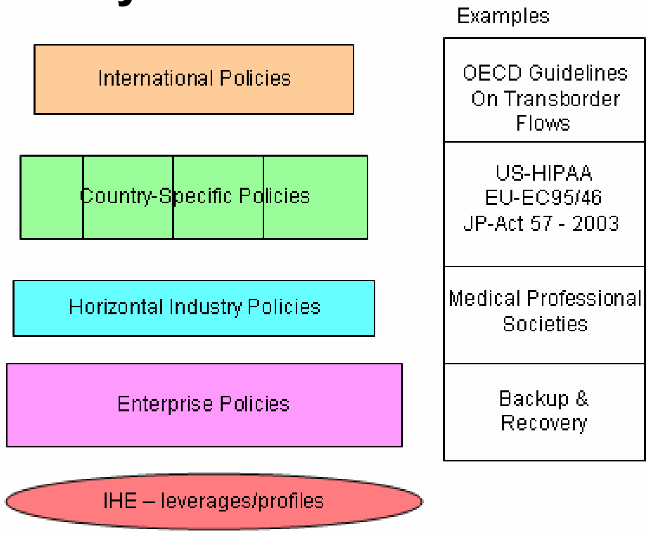
caBIG™
cancer Biomedical Informatics Grid™

The caBIG Trust Fabric

Security Basics

- **Security is conceptually built on layers in at least three spaces**
  - policy
  - procedure
  - technical implementation

- **Example: IHE Defined Policy Environment**



Examples

International Policies

OECD Guidelines On Transborder Flows

Country-Specific Policies

US-HIPAA
EU-EC95/46
JP-Act 57 - 2003

Horizontal Industry Policies

Medical Professional Societies

Enterprise Policies

Backup & Recovery

IHE – leverages/profiles

Material derived from the IHE IT Infrastructure White Paper, 2007 draft, available at http://www.ihe.net

caBIG Trust Fabric

- **What is the Trust Fabric?**
  - caGrid Security Framework – Technical Overview
  - Security Framework Components
- **Mapping the Technical Framework to the Policy:** caGrid Security Service Compliance with NIST E-Authentication LOA 2
- **Implementing the Trust Fabric**
  - Status of Implementation: **Developing caBIG Security Policies and Procedures**

Current Status

- Basic technical Infrastructure in place – caGrid GAARDS infrastructure
- Basic decision to leverage extensive, and growing, OMB/NIST E-Authentication federated infrastructure
  - Supports federation initiatives and multi-institutional workflows
  - Defines four “Levels of Assurance” around individual identity
- Security Working Group focused on developing policies and procedures for minimal to moderately sensitive data
  - Authentication/identity management issues
  - Authorization/privilege management issues

Authentication: Levels of Assurance

Federal E-Authentication Initiative

http://www.cio.gov/eauthentication/

- Levels of assurance (Different Requirements)
  - Level 1 – e.g., no identity vetting
  - Level 2 – e.g., specific identity vetting requirements
  - Level 3 – e.g., cryptographic tokens required
  - Level 4 – e.g., cryptographic hard tokens required
- Credential Assessment Framework Suite (CAF)
- Federal Bridge Certification Authority (FBCA)
  - http://www.cio.gov/fbca/
  - The FBCA is an information system that facilitates an entity accepting certificates issued by another entity for a transaction.

Authorization

Process of determining if an *authenticated* person qualifies to conduct certain activities in cyberspace.

1. A relying party obtains certain attributes from a source of authority (SOA) to determine if an authenticated claimant qualifies for certain privileges.
2. The SOA “knows” the certified identifier of an authenticated person.
3. The SOA verifies that the identified person has certain attributes.
4. Upon verification of the required personal attributes, the relying parting authorizes the authenticated claimant to conduct the desired activities.

Authorization Polices and Procedures

A community of relying parties that has a common operational context requiring “trusted” authorization of individual activities must agreed upon policies and procedures that define:

- ***Designated Sources of Authorities (SOAs) and Sources of Records (SORs) for personal attributes.***
- ***Specific personal attributes and allowed values required for authorization decisions.***
- ***What constitutes an acceptable certified identifier of physical person.***
- ***How physical identity is reconciled when SOAs use different identifiers for the same individual.***
- ***How attribute values are managed – i.e. assigned, verified and altered in a timely fashion by an SOR.***

Selecting Appropriate AuthN/AuthZ Controls

- **Authentication - Levels of Assurance (LOA)**
  - *LOA provides some idea of the risk that a claimant other than the certified physical person may have authenticated to a system.*
    - ✓ **How thoroughly was physical identity vetted by the credentialing authority?**
    - ✓ **Was the activator (e.g. password) of the credential actually given to the vetted person?**
    - ✓ **How easy is it for others to use the credential of someone else?**
- **Authorization – Areas of Risk**
  - ***Specified and validated SOA, SOR and privilege management processes provide some idea of the risk that inappropriate authorizations may be granted to correctly authenticated individuals.***
    - ***Do all relying parties use the same identifier for the same physical person? If no, how good is the identity process?***
    - ***Do all SOAs and SORs use standard processes and well defined attribute sets and values?***
    - ***What level of authentication is required to assign attribute values or system privileges.***

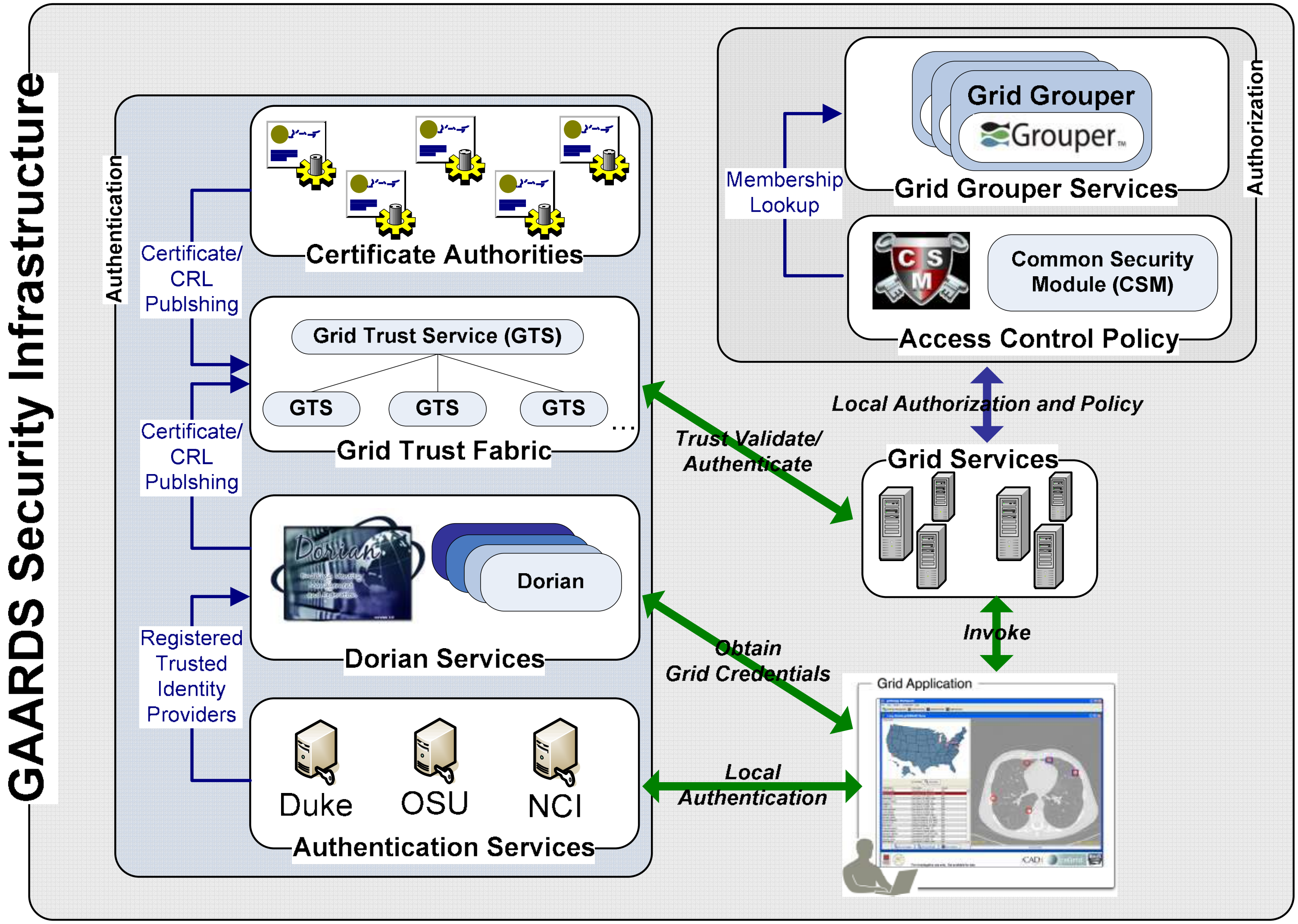
caBIG™ Security Framework Overview

- **Need for Secure Data Exchange:**
  - In the cancer research community, the assurance of protection and privacy of patient related and sensitive information and the protection of intellectual property and are critical to the success of interoperable exchange of research data.
- **Security Solution:**
  - The caBIG™ Security Framework provides services and tools for the administration and enforcement of security policy in the enterprise caGrid. These services and tools are coupled with policies and procedures for federated identity management across various security Levels of Assurance.

Security Framework Components

- The caGrid Security Infrastructure services and tools that enable secure data exchange include:
  - **Dorian**–A grid service for the provisioning and management of grid users accounts
  - **Grid Trust Service (GTS)**–A grid-wide mechanism for maintaining and provisioning a federated trust fabric consisting of trusted certificate authorities
  - **Grid Grouper**–A service that provides a group-based authorization solution for the Grid
  - **Credential Delegation Service (CDS)**–Enables users/services (delegator) to delegate their Grid credentials to other users/services (delegatee).
  - **Web Single Sign On (WebSSO)**–Provides a comprehensive, Single Sign On (SSO) solution for web applications using GAARDS
  - **Authentication Service**–Provides a framework for issuing SAML assertions for existing credential providers such that they may easily integrated with Dorian and other grid credential providers
  - **Common Security Module (CSM)**–Provides a centralized approach to managing and enforcing access control policy authorization

caGrid Security Infrastructure



Building out the caBIG Trust Fabric

- The DSIC Workspace and the Security Working Group are optimizing the existing trust fabric by developing security polices and procedures that will scale to fit the needs of the cancer research community and serve as a model platform for other research.

- The caGrid serves as an infrastructure and procedural model for other research collaborations (NIH/NHLBI's CVRG, UK NCRI's ONIX) and for health data exchanges such as the HHS NHIN initiative.

- The DSIC WS is building out tools to facilitate assessing the levels of data access and security for various types of data (PHI, deidentified, LDS), to enable automated construction of many contractual agreements between providers and users of data and to inform key individuals in institutions about the caBIG program and caGrid processes for securing data access.

Mapping the Technical Framework to the Policy - *NIST Level 2 Assurance*

Level of Assurance	Data Sensitivity Level	caBIG™ Security Policies/Procedures
LOA 1	Low sensitivity data (no IP value, no IRB or sponsor restrictions)	• caGrid Level 1 Host Trust Agreement for Interfederation (to bring new identity providers into the trust federation) • LOA1 caGrid Certificate Policy and Practice Statement
LOA 2	Moderate sensitivity data (deidentified data, Limited Data Sets, moderate IP value and moderate institutional/IRB/ sponsor restrictions)	• LOA2 caGrid Certificate Policy and Practice Statement • Underway: LOA2 caGrid Level 2 Host Trust Agreement and pilot Technical Implementation at an academic institution (UT)

Mapping the Technical Framework to the Policy - *NIST Level 2 Assurance*

Level of Assurance	Data Sensitivity Level	caBIG™ Security Policies/Procedures
LOA 1	Low sensitivity data (no IP value, no IRB or sponsor restrictions)	• caGrid Level 1 Host Trust Agreement for Interfederation (to bring new identity providers into the trust federation) • LOA1 caGrid Certificate Policy and Practice Statement
LOA 2	Moderate sensitivity data (moderate IP value and institutional/IRB/ sponsor restrictions)	• LOA2 caGrid Certificate Policy and Practice Statement •**Underway: LOA2 Technical Implementation at academi institution**

To support LOA2 implementation with caGrid at a test institution, an interface is currently under development to enable the Dorian identity provider to issue Grid credentials through federated security assertions.

Security Policies and Procedures for Non-sensitive Data

NCI/caGrid Host Agreement – Level 1 Data:

- **Required for hosting Level 1 data services; intended for sharing non-sensitive data (de-identified or non-human)**
- **Can be signed by individuals who host Grid services; they determine if additional review/signature is required**
- **Purpose: describe information security responsibilities of Grid Hosts**
- **Grid Host is responsible for:**
  - Complying with applicable laws and regulations
  - Implementing policies and procedures to enable compliance with caBIG™ security principles
  - Reporting security breaches and participating in security investigations
  - Applying system upgrades and patches
  - Maintaining security of host certificates

Security Policies and Procedures for Low to Medium Sensitivity Data

- **Governance Model for Authentication for services available to persons who authenticate at LOA 2**
    - caBIG™ Identity Provider Federation Policy document (to describe governance model)
    - Model Trust Agreement for Interfederation (to bring new identity providers into the trust federation)
    - LOA 2 Technical Implementation (based on understanding of policies and issues related to implementation identified by Security Working Group)

Inter-Federation Project

- **NIH and the University of Texas Identity Management (IdM) Federation will develop an inter-federation agreement at LOA 2.**
  - ✓ ***NIH will rely on the GSA E-Authentication Credential Assessment Framework (CAF) as the standard for determining whether the UT Federation SAML assertions meet OMB/NIST LOA 2.***
  - ✓ ***NIH and UT will agree how and by whom the CAF assessment will be performed.***
  - □ ***Upon successful completion of an NIH/UT agreement, efforts will be made to develop similar inter-federation agreements with other federations such as InCommon.***

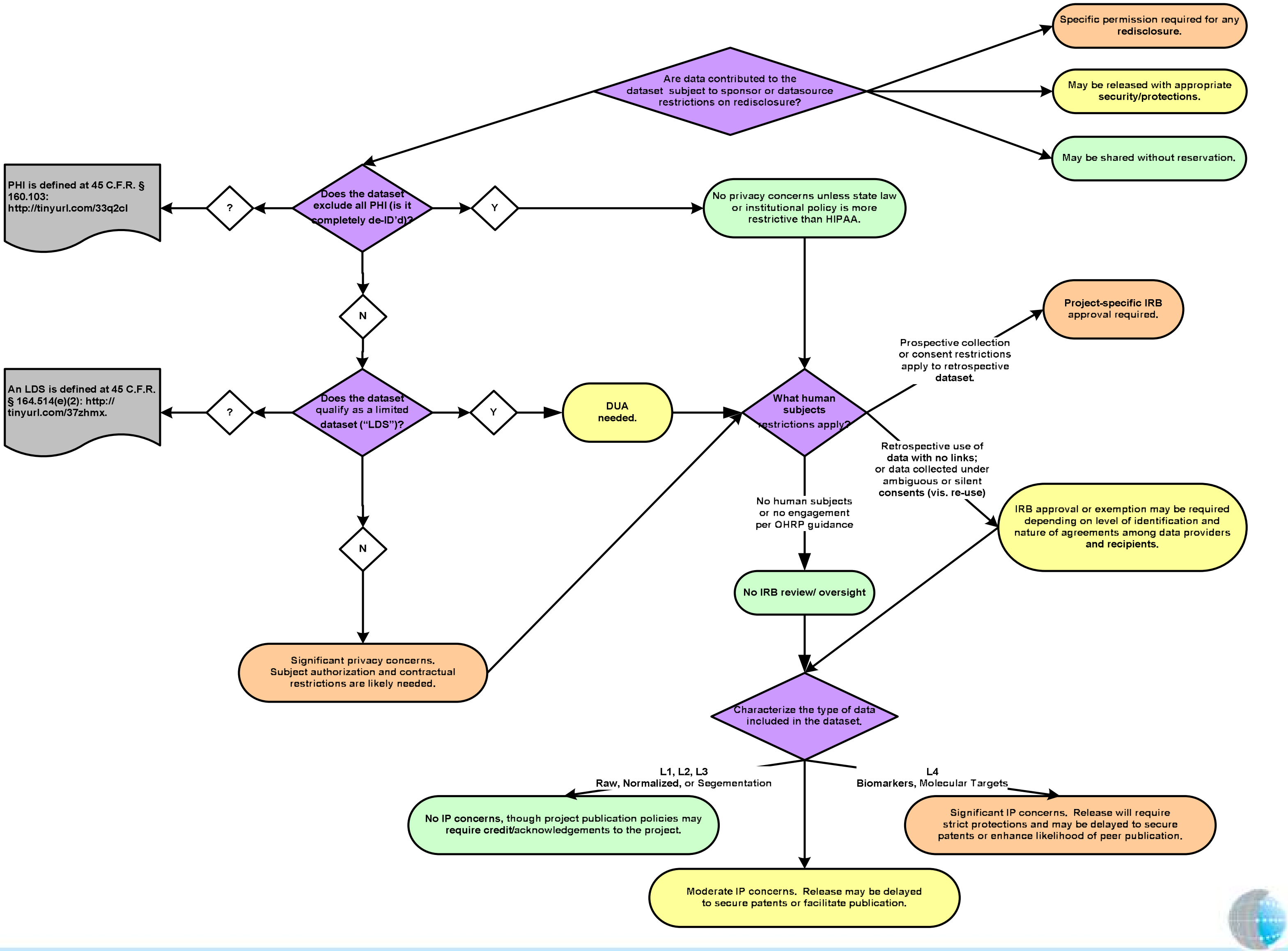
- **Develop and test an open-source interface for sending SAML assertions between and an institutional Shibboleth identity provider (IdP) and DORIAN**

National Cancer Institute

caBIG™
cancer Biomedical Informatics Grid™

DSSF Decision Support Tools

DSSF Decision Support Tool



Are data contributed to the dataset subject to sponsor or datasource restrictions on redisclosure?

Specific permission required for any redisclosure.

May be released with appropriate security/protections.

May be shared without reservation.

PHI is defined at 45 C.F.R. § 160.103: http://tinyurl.com/33q2cl

?

Does the dataset exclude all PHI (is it completely de-ID'd)?

Y

No privacy concerns unless state law or institutional policy is more restrictive than HIPAA.

N

An LDS is defined at 45 C.F.R. § 164.514(e)(2): http://tinyurl.com/37zhmx.

?

Does the dataset qualify as a limited dataset ("LDS")?

Y

DUA needed.

N

Significant privacy concerns. Subject authorization and contractual restrictions are likely needed.

What human subjects restrictions apply?

Prospective collection or consent restrictions apply to retrospective dataset.

Project-specific IRB approval required.

Retrospective use of data with no links; or data collected under ambiguous or silent consents (vis. re-use)

IRB approval or exemption may be required depending on level of identification and nature of agreements among data providers and recipients.

No human subjects or no engagement per OHRP guidance

No IRB review/ oversight

Characterize the type of data included in the dataset.

L1, L2, L3
Raw, Normalized, or Segementation

No IP concerns, though project publication policies may require credit/acknowledgements to the project.

L4
Biomarkers, Molecular Targets

Significant IP concerns. Release will require strict protections and may be delayed to secure patents or enhance likelihood of peer publication.

Moderate IP concerns. Release may be delayed to secure patents or facilitate publication.

Developers & Audience

Issue	Likely Subject Matter Expert(s)
Privacy/Confidentiality	Privacy Officers Compliance Officers CIOs Legal Counsel
Research Policies/Ethics	HRPP Directors IRB Staff IRB Chairs Legal Counsel
Proprietary Issues (Institution/Researcher)	Tech Transfer Officials Research Associate Deans Legal Counsel
Sponsor Requirements	Sponsored Programs Offices Legal Counsel

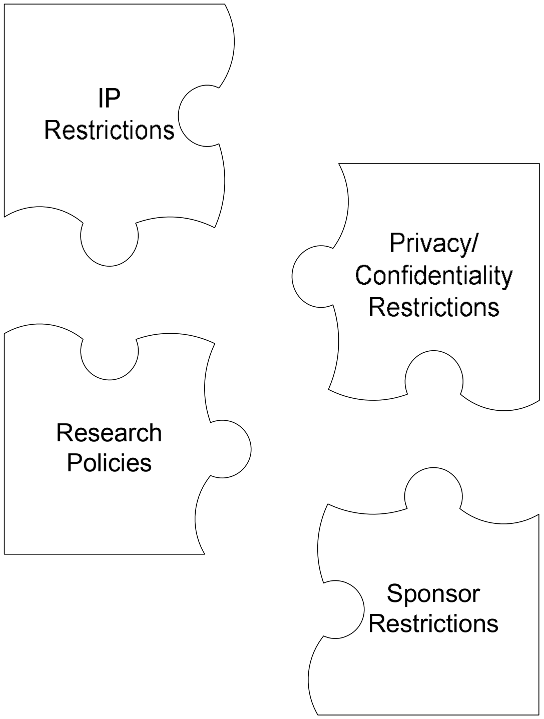
Usual Disclaimers

- ***The analyses represented in the following slides reflect current group understanding of applicable federal laws and regulations, but do not reflect more restrictive state laws or institutional policies and are no substitute for legal advice from your own institutional attorneys.***

- ***Questions, comments, and suggestions are always welcome. These tools are continuously improved with contributions from workspace contributors.***

Executive Summary

- **Four separate analyses contribute to an understanding of the nature of any mandatory legal restrictions on data sharing and agreements, if any, necessary to facilitate proposed exchanges.**
- **The analyses are conducted locally to assure institutional compliance with state law and institutional policy.**
- **Future state:**
    - Any necessary agreements are incorporated into the various applications to facilitate rapid data exchange and eliminate the need for bilateral or multilateral written agreements in most cases.
    - Security framework supports authentication and authorization needs consistent with HIPAA, Common Rule/FDA confidentiality requirements, and industry standards.


IP Restrictions

Privacy/ Confidentiality Restrictions

Research Policies

Sponsor Restrictions

Privacy Restrictions

- **Do federal or state laws, or your institution's privacy or confidentiality policies, restrict disclosure (research or IRB policies are addressed separately under "ethics")?**

- Questions/Issues
  - **Are data to be shared completely de-identified (per HIPAA definition)? Do they qualify as a "limited data set"?**
  - **Are the data otherwise identifiable (linkable) to individuals (e.g., SNP data where there is some reference dataset reasonably available to recipients)?**
  - **Do state laws further restrict disclosure?**
  - **Do institutional privacy/confidentiality policies further restrict disclosure?**
  - **Are there any mandates to disclose (e.g., funding agencies, ICMJE, www.clinicaltrials.gov, BMT)**

De-Identified Data

HIPAA permits a deidentified data set (“DDS”) – one that omits all direct and indirect identifiers – to be shared with researchers without restriction. To be considered de-identified, the data set must exclude the following elements with respect to an individual or the individual’s relatives, employers, or household members:

- **Name**
- **Address, city, and other geographic information smaller than state (3-digit zip code may be included only for an area where more than 20,000 people live)**
- **All elements of date (except year), plus age and any date, including year, if age is over 89 (date or age ranges may be included)**
- **Telephone, fax, e-mail, web URL, IP address**
- **Social security number, medical record number, health plan beneficiary number, account number, certificate or license number**
- **Vehicle identifier (e.g., license plate or serial number)**
- **Device identifier (e.g., serial number)**
- **Biometric identifier (e.g., finger print or voice print – DNA is *not* considered a direct or indirect identifier under HIPAA)**
- **Full-face photograph or comparable image**
- **Any other unique identifying number, characteristic, or code (except a code used for linking purposes as prescribed by HIPAA).**

Privacy/
Confidentiality
Restrictions

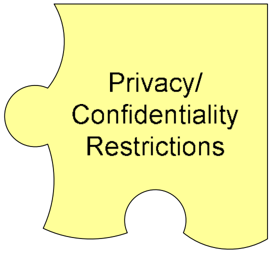
Alternatively, a dataset is de-identified if an appropriately qualified statistician determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated receipient to identify an individual who is the subject of the information; and documents the methods and results of the analysis.

NOTE: Even if deidentified as provided above, if the covered entity has actual knowledge that the dataset could be used alone or with other information to identify an individual who is the subject of the information, the data are not considered to be deidentified. In addition, state laws or institutional policy may further restrict disclosure.

Limited Data Set

- **A limited data set ("LDS") is similar to a de-identified data set but may include geographic information other than street address; dates and ages; and other unique identifying numbers, characteristics, or codes.**
- **An LDS may be shared with researchers who sign a *data use agreement* to assure that they will:**
    - Use the data only for the designated research
    - Protect the data against inappropriate disclosure
    - Not use the information to re-identify the included individuals.
- **HIPAA prescribes very specifically the requirements for a data use agreement. An agreement that meets HIPAA's standards is under development by DSIC.**

NOTE: State laws or institutional privacy policies may further restrict use or disclosure.



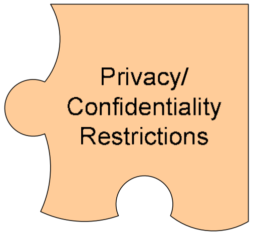
Privacy/
Confidentiality
Restrictions

Identifiable Data

- **HIPAA generally permits an identifiable data set to be shared only with the specific written agreement (“authorization”) of the individuals whose data will be disclosed, or under a waiver of authorization approved by an IRB or privacy board. [Preliminary data reviews “preparatory to research,” and research involving information of only deceased subjects are also permitted with appropriate certifications but typically are not relevant in this context.]**

- **Even with written authorization, most institutions will share identifiable data only under agreements that assure confidential treatment of the data to be exchanged. These agreements may be simple or more complex, depending on other considerations.**

- **DSIC’s work on development of agreements and processes to facilitate data sharing is discussed below.**

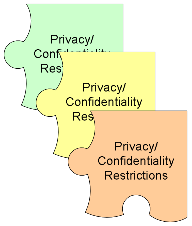


Privacy/ Confidentiality Restrictions

Additional Privacy Considerations

- **Do state laws further restrict disclosure?**
  - Some states have enacted laws considered to be “more restrictive” than HIPAA, typically to provide additional protection to particularly “sensitive” data:
    - Genetic testing
    - Cancer diagnosis
    - HIV/AIDS and other serious communicable disease information
    - Substance abuse or mental health treatment

    Many of these laws include exceptions for research uses of the data or for data sets that have been de-identified. Some may require contracts supporting exchange of sensitive data to include certain language.
  - Because institutions determine at the local level whether to share any particular data set, they are responsible for interpreting these state laws and are not placed in the position of relying on others who might have different interpretations.
- **Do institutional policies further restrict disclosure?**
  - Research institutions rarely restrict data exchange beyond the requirements of federal or state law for privacy reasons (analysis of other elements of the framework is included in the other sections of this presentation and other DSSF materials).
- **Are there mandates for disclosure (e.g., from funding agencies, ICMJE/FDAAA, public health registries, etc.)?**

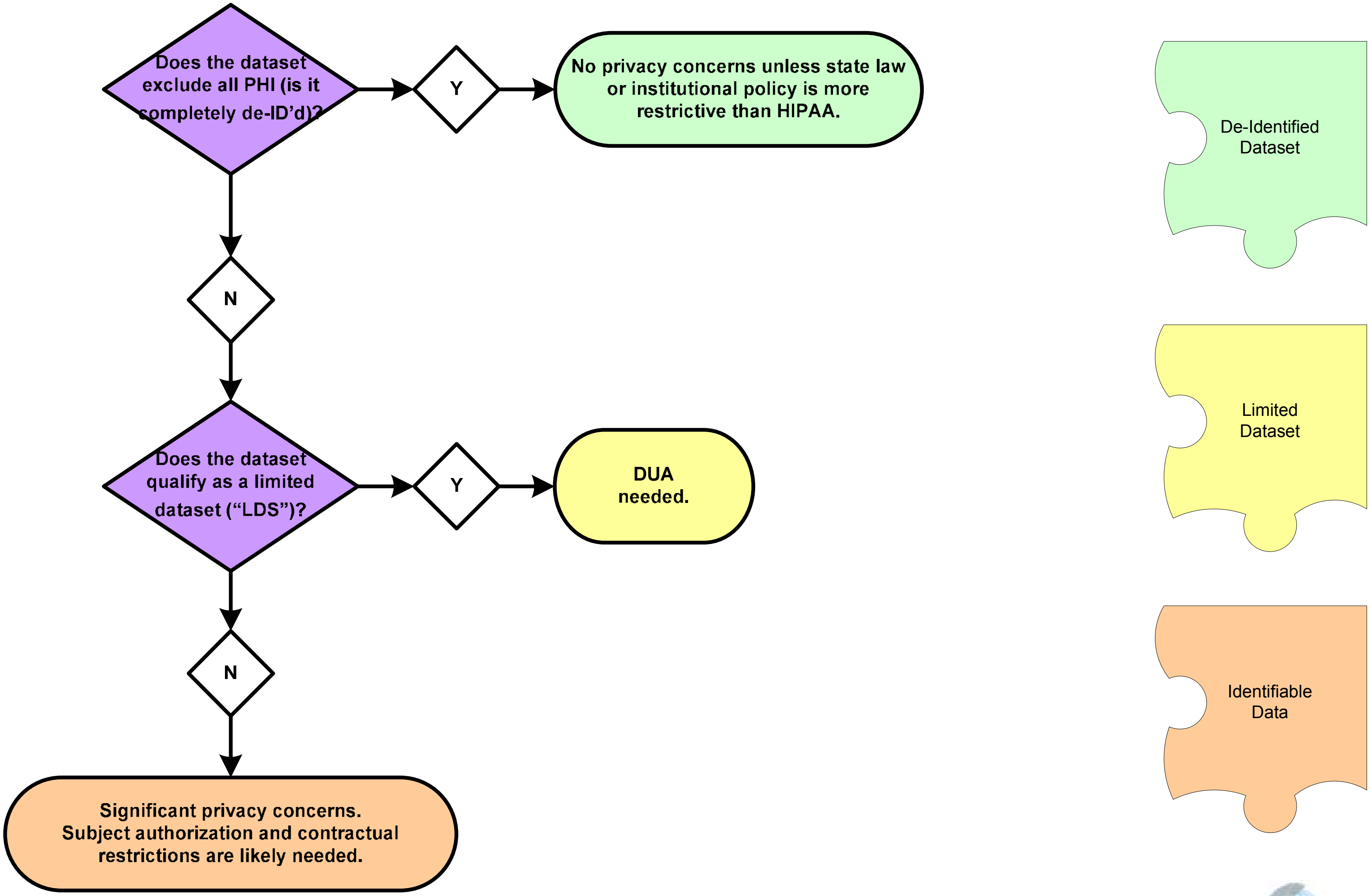


Overcoming Privacy Barriers

1. **Level of Identification**
   - Does the dataset to be shared include identifiable information?
   - If so, can identifiers be removed to create a “limited data set” or a “deidentified data set” without compromising the integrity of the research (*note: a deidentified data set may include links or codes to facilitate reidentification*)?
   - If identifiers cannot be removed, does disclosure meet another HIPAA exception
     - Review of decedents’ information
     - Review preparatory to research (no data off-site -> inapplicable)
     - Waiver of authorization (granted by an IRB or privacy board)
2. **Protective Agreements: even if HIPAA (or applicable state law or institutional policy) restricts disclosure, restrictions generally can be addressed through use of appropriate agreements**
   - Deidentified data set: none is generally necessary
   - Limited data set: data use agreement
   - Identifiable data: restrictive confidentiality agreements (not necessarily required from a federal regulatory perspective with subject authorization or if an authorization exception applies but practically important to assure subject protections and as industry “best practice”)

Note: State laws and institutional policies can significantly impact this analysis. Many states impose special protections on genetic information, cancer information, behavioral health records, etc. Knowledge of these laws is essential to accurately identify privacy barriers and evaluate how best to overcome them.

Privacy Summary



Does the dataset exclude all PHI (is it completely de-ID'd)?

Y

No privacy concerns unless state law or institutional policy is more restrictive than HIPAA.

N

Does the dataset qualify as a limited dataset ("LDS")?

Y

DUA needed.

N

Significant privacy concerns. Subject authorization and contractual restrictions are likely needed.

De-Identified Dataset

Limited Dataset

Identifiable Data

IRB/Ethical Restrictions

- **Do the Common Rule, FDA regulations, or your institutional research or IRB policies restrict the proposed disclosure, or were the data collected under an informed consent document (or process) that would prohibit the disclosure?**
- Questions/Issues
  - **Is the project "human subjects research" or a "clinical investigation" under the Common Rule, FDA, or local institutional policies? Is the data provider "engaged" in the research?**
  - **Is the research potentially eligible for an exemption from continuing IRB oversight; does it involve "secondary" use of data originally collected under consents that approved re-use or were silent or ambiguous?**
  - **What were the circumstances of the original data collection (purpose, consent documents, etc.)?**
  - **Are there any explicit restrictions on data sharing?**
    - **Would disclosure be inconsistent with protocol or policies under which data originally were collected?**
    - **Would disclosure be inconsistent with the original consent (or IRB-approved waiver of consent), or has consent been withdrawn?**

Definitions (Common Rule)

A human subject is a living individual about whom an investigator conducting research obtains data through interaction (e.g., survey) or intervention (e.g., venipuncture or experimental treatment) with the individual, or identifiable private information.

An investigator is someone involved in the design, analysis, or publication of results. OHRP does not necessarily consider the act of furnishing identifiable or coded private information or specimens to an investigator to, in and of itself, constitute research.

Obtaining means receiving or accessing identifiable private information or identifiable specimens for research purposes. (Obtaining includes study or analysis of data or specimens already in the investigator’s possession.)

Usage of Terms

- **“Data” includes written information, images, specimens, etc.**
- **When individual identifiers are removed from data sets, the resulting information may be referred to as “anonymized,” “de-identified,” “coded,” or some similar moniker.**
- **For purposes of this presentation:**
  - “Anonymized” means that data cannot by any means be linked to specific individuals
  - “De-identified” means certain identifying elements have been removed so that the data are no longer considered “protected health information” under HIPAA. 45 C.F.R. § 164.514(b).
  - “Coded” means that directly identifying information has been removed but a code has been retained to permit future re-identification.
- **Coded information may or may not qualify as “de-identified.”**
- **De-identified data may, consistent with HIPAA, be coded for future re-identification if the code is unrelated to any of the direct identifiers referenced in HIPAA, is used only for re-identification purposes, and is otherwise secured.**

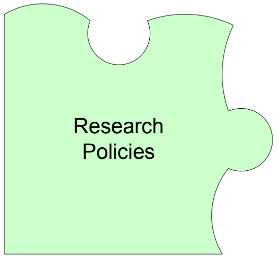
No Human “Subjects”

- **Data to be shared relate solely to deceased individuals (individuals must be living to be considered “human subjects”)**

- **Those disclosing the data are not involved in the design, analysis, or publication of results of the current project (but data recipients may be engaged) – more on this below**

- **Data to be shared are completely de-identified before the start of the study and consent forms (or IRB-approved waiver) under which data originally were collected did not specifically restrict or limit use of de-identified data for the current project or future research generally.**

NOTE: if human specimens are to be exchanged, see further analysis below.



12/2006 OHRP Draft Engagement Guidance

- **Unless it receives a direct HHS award for conducting the research (even if all of the research activities are subcontracted out), a data/specimen provider institution is not “engaged” in the research if:**
  - The data/specimens to be disclosed or transferred (which may be identifiable) originally were collected for purposes other than the current project (e.g., clinical care or an unrelated research study); *and*
  - Disclosure is not inconsistent with original consents (or IRB-approved waiver).

Note: the recipient institution is engaged in the research, at least if it receives identifiable data and, accordingly, must secure IRB approval and waiver of informed consent before proceeding with the study.

Coded Data
[8/2004 OHRP “Coded” Guidance]

- **A project is not considered human subjects research if:**
  - Data originally were collected for purposes other than the current project; and
  - Data are “coded;” and
  - Investigators performing the research can’t readily ascertain the identity of the affected individuals (e.g., based on contractual terms agreed to by the recipients, or based on existing institutional or IRB policies prohibiting release of keys to recipients; and
  - Data providers are not involved in the design, analysis, or publication of results of the current project; and
  - No conflicts with original consents

Research Policies

Exempt Research or “Master Protocol”

Research is exempt from IRB oversight under the Common Rule if:

- The data to be used already existed at the time the study started; and
- The data to be used are not directly or indirectly linked to individual subjects; and
- The IRB or other designated institutional office/official has approved the exemption

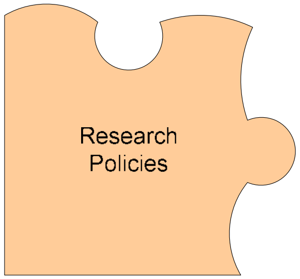
No new IRB approval or informed consent (or waiver) is required if:

- The project is proceeding under a “master” or “umbrella” protocol that covers a broad range of activities or multiple sub-studies
- The project is performed consistent with that protocol.

Research Policies

Other Projects

- **Prospective collection of data and/or specimens**
  - Prospective IRB approval is required
  - Informed consent (or IRB-approved waiver of consent) is required
- **In vitro diagnostic studies**
  - Prospective IRB approval is required
  - Informed consent is not required if the specimens are de-identified per recent FDA guidance
- “**Secondary” research inconsistent with original consent**
  - Depending on local policies and ethical considerations, data/specimens may not be used, or possibly may be used with prior IRB approval and waiver of new consent
- **Consent withdrawn**
  - There is some controversy about the meaning of the term “withdrawal” and different consent documents treat the issue differently.
  - Examples of consent provisions related to withdrawals:
    - If a subject withdraws from a study, previously collected information will be de-identified and may be used to continue the project or for other appropriate purposes
    - Promise that data and specimens will be destroyed upon withdrawal.



Research Policies

Overcoming Research Policy Barriers

IRB Oversight

1. **Distinguish human subjects research from unregulated research**
   - Human subjects are alive
   - OHRP has issued guidance on “coded information and specimens” (http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm) and draft guidance on institutional “engagement” (http://www.hhs.gov/ohrp/requests/engage.html) that together define many activities currently regulated by IRBs as non-human subjects research
   - In most institutions, non-human subjects research is not subject to IRB oversight, though who makes that decision with respect to any given project varies

2. **Determine whether the research is “exempt” or whether a proposed project or inquiry is covered under a “master” or “umbrella” protocol**
   - Studies involving previously collected data that cannot be directly or indirectly linked to living individuals are eligible for exemption, which generally must be granted by the IRB
   - IRBs may sometimes approve “master” or “umbrella” protocols that cover a broad range of individual projects or analyses

3. **For non-exempt human subjects research, consider alternatives to multi-institutional approval**
   - CIRB
   - Commercial IRB
   - Defer or accept review under an IRB Authorization Agreement (http://www.hhs.gov/ohrp/humansubjects/assurance/iprotsup.rtf)

Overcoming IRB Barriers

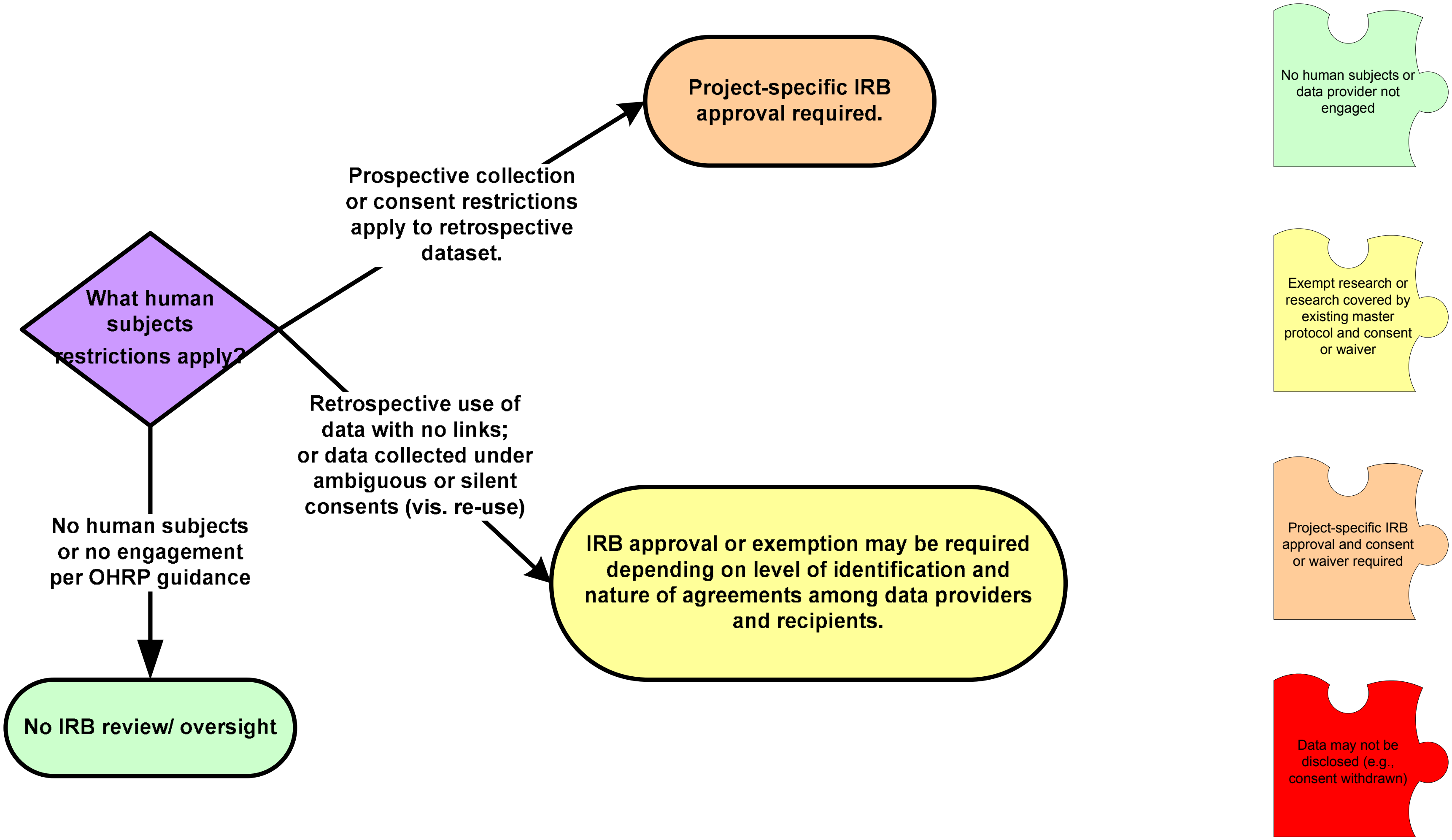
Informed Consent

1. **Explicit permission to share data with researchers via caGrid (or more broadly through a data/specimen registry) can help eliminate IRB or broader ethical barriers**

2. **Absent explicit permission, IRBs may permit retrospective research on data or specimens previously collected under clinical or research consent documents that were silent or ambiguous about future use for unspecified analyses conducted by the original research team or others**
   - Many older consents include explicit language that restricts use of data or specimens to the current project
   - Response varies: IRBs may permit reuse under a waiver if deemed consistent with original intent of the consent, may permit re-contact with subjects to solicit explicit permission, or may bar reuse and re-contact

3. **Explicit permission (or IRB-approved waiver) is required for prospective collection of data or specimens**
   - Waiver may be difficult to secure because it requires a showing of “impracticability”

Research Policies Summary



What human subjects restrictions apply?

Prospective collection or consent restrictions apply to retrospective dataset.

Project-specific IRB approval required.

Retrospective use of data with no links; or data collected under ambiguous or silent consents (vis. re-use)

IRB approval or exemption may be required depending on level of identification and nature of agreements among data providers and recipients.

No human subjects or no engagement per OHRP guidance

No IRB review/ oversight

🟩 No human subjects or data provider not engaged

🟨 Exempt research or research covered by existing master protocol and consent or waiver

🟧 Project-specific IRB approval and consent or waiver required

🟥 Data may not be disclosed (e.g., consent withdrawn)

Proprietary Restrictions

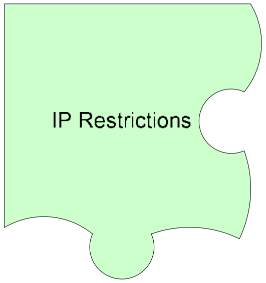
- **Does the need for protection from an institutional and/or investigator perspective restrict the proposed disclosure of research data?**

- Questions/Issues:
  - **What is the intellectual property status of the data? Is it valueless from an IP perspective, is a patent intended but not yet filed or published, or is it already patented?**
  - **What is the publication status? Are the results already in the public domain? Are they published but not publicly accessible? Do they remain unpublished? Is additional work needed to assure full and accurate analysis by anticipated users?**
  - **Does the data owner or steward have other reasons for restricting dissemination?**

“Low Sensitivity Data”

- **Intellectual property status**
  - The data do not disclose a potentially patentable invention or, if they do, no patent application is intended to be filed on such an invention
  - The data disclose a patentable invention but a patent application has been filed and already published in the literature or through the patent office
  - Data have no intrinsic commercial value, *i.e.,* no academic or commercial entity is likely to pay fees for access to the data.

- **Publication status**
  - Data have been published in the scientific literature or on the web, or deposited into public repository.

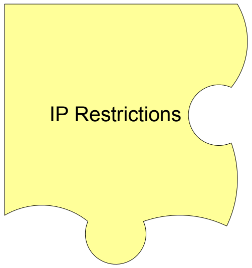


“Moderate Sensitivity Data”

- **Intellectual property status**
  - Data disclose a potentially patentable invention on which the owner intends to file a patent application.
  - Patent application (provisional or other) already filed on the data but data not published by the patent office
  - Data may have intrinsic value that requires protection, e.g., data that is time consuming or expensive to replicate, that can be realized through licensing

- **Publication status**
  - Results based on data are not yet submitted for publication
  - Publications based on the data have been submitted but not yet published, or published with restrictions imposed by the copyright holder
  - Data currently are only available within a consortium or other limited group under mutual confidentiality obligations

- **Miscellaneous**
  - Other considerations render the data somewhat sensitive from the institution’s or researcher’s perspective

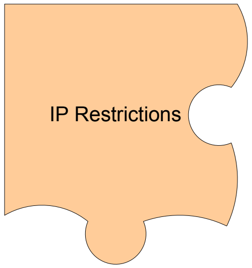


“High Sensitivity Data”

- **Intellectual property status**
  - Data have intrinsic commercial value to both academic and commercial entities that can be realized through licensing (active expression of interest from both academic and commercial entities)

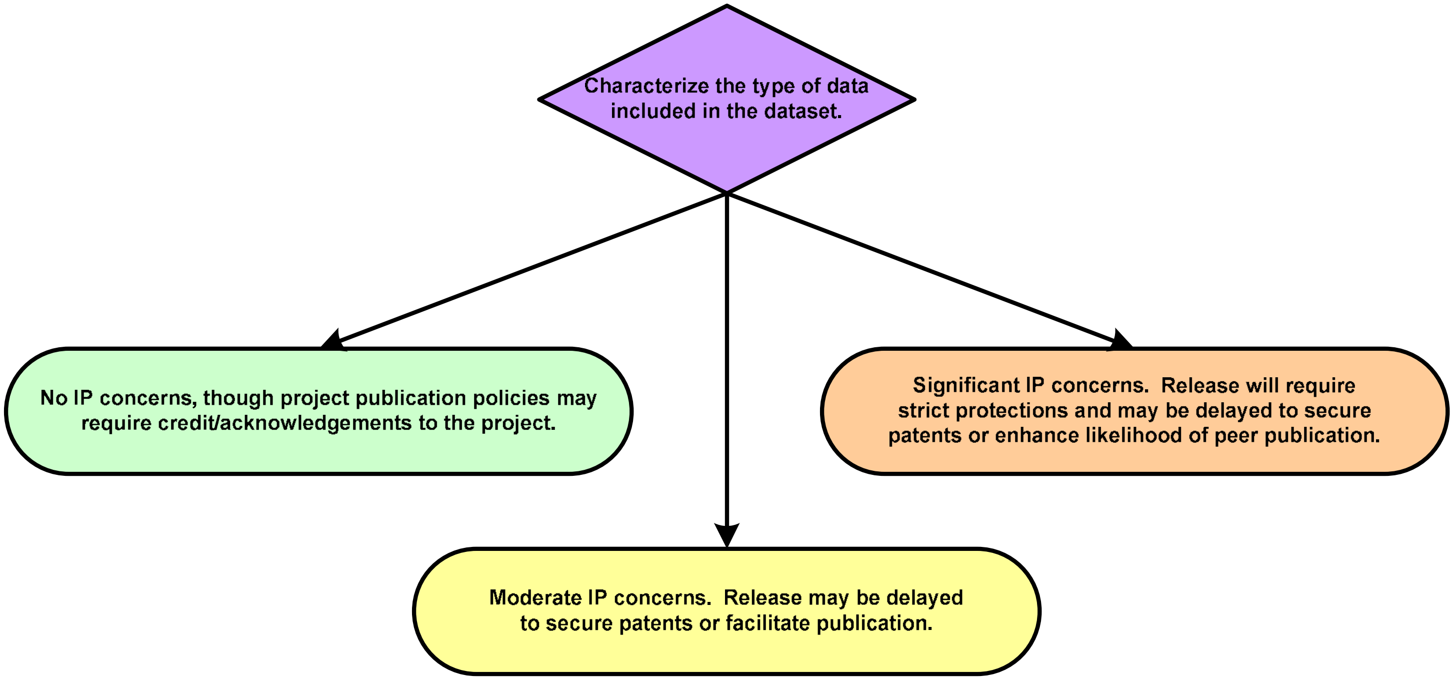
- **Publication status**
  - Results based on the data are unpublished
  - Data cannot be fully analyzed without additional data to be generated in the future

- **Miscellaneous**
  - Other considerations render the data somewhat sensitive from the institution’s or researcher’s perspective



IP Restrictions

IP Restrictions Summary

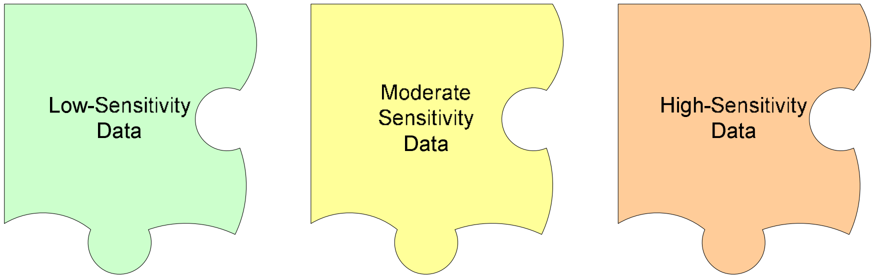


Characterize the type of data included in the dataset.

No IP concerns, though project publication policies may require credit/acknowledgements to the project.

Moderate IP concerns. Release may be delayed to secure patents or facilitate publication.

Significant IP concerns. Release will require strict protections and may be delayed to secure patents or enhance likelihood of peer publication.



Low-Sensitivity Data

Moderate Sensitivity Data

High-Sensitivity Data

Sponsor Requirements

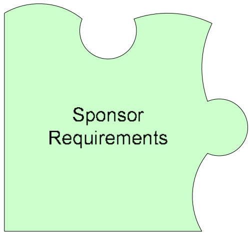
- **Do the terms and conditions in any funding grants or contracts (from the government or private sources), or other agreements, prohibit or restrict disclosure?**

- Questions/Issues:
  - **Do the terms of any agreements with sponsors governing the original data collection or creation delay or otherwise limit, restrict, or prohibit disclosure/**
  - **Do the terms of any agreements with original data sources delay or otherwise limit, restrict, or prohibit disclosure?**

“Low Sensitivity Data”

Funding or other related agreements contain no restrictions or require only attribution

- The institution and investigator are not bound by any agreements that restrict the right to share data with others including those outside the institution
- The institution or investigator is bound by an agreement requiring only attribution of the source of the data to be disseminated

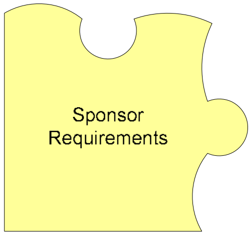


Sponsor Requirements

“Moderate Sensitivity Data”

Funding or other agreement includes any of the following types of provisions:

- Imposes restrictions on data sharing for a limited period of time (e.g., sharing only after publication by sponsor, or only after project participants have an exclusive time period to review the data or only after a related patient application has been filed)
- Allows data sharing only with non-profit entities or other defined groups
- Allows data sharing only for non-commercial or other restricted purposes
- Allows use but not dissemination of data derived from data provided by the disclosing institution or investigator, or funded by the applicable sponsor

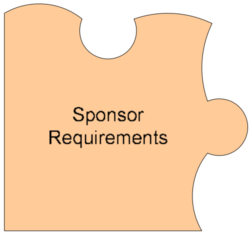


Sponsor Requirements

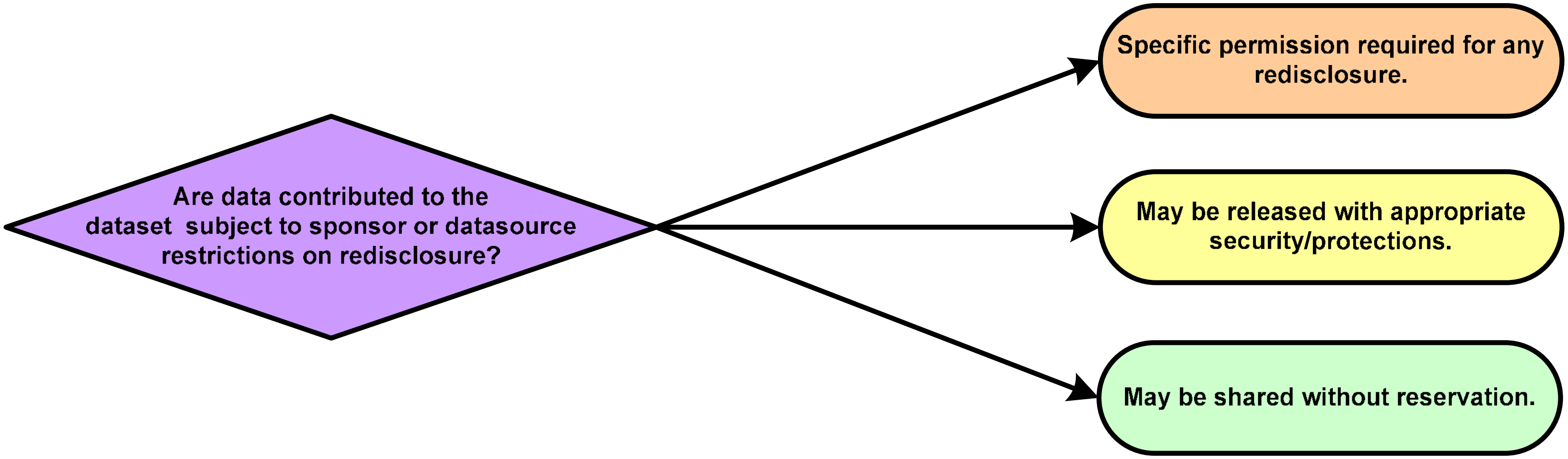
“High Sensitivity Data”

Funding or other agreement:

- Allows transfer only under defined security conditions
- Requires the receiving or funded entity to disclose, license, or assign results derived from the data to the sponsor or another specified party
- Otherwise restricts receiving institutions’ ability to retransfer data



Sponsor Requirements Summary

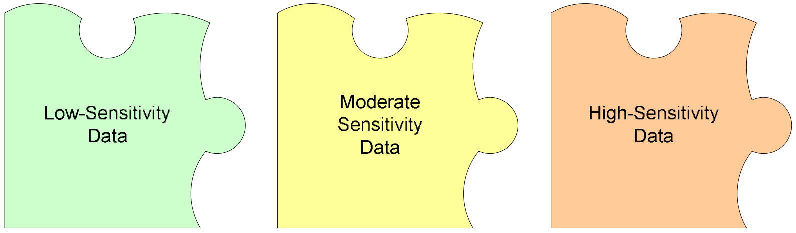


Are data contributed to the dataset subject to sponsor or datasource restrictions on redisclosure?

Specific permission required for any redisclosure.

May be released with appropriate security/protections.

May be shared without reservation.

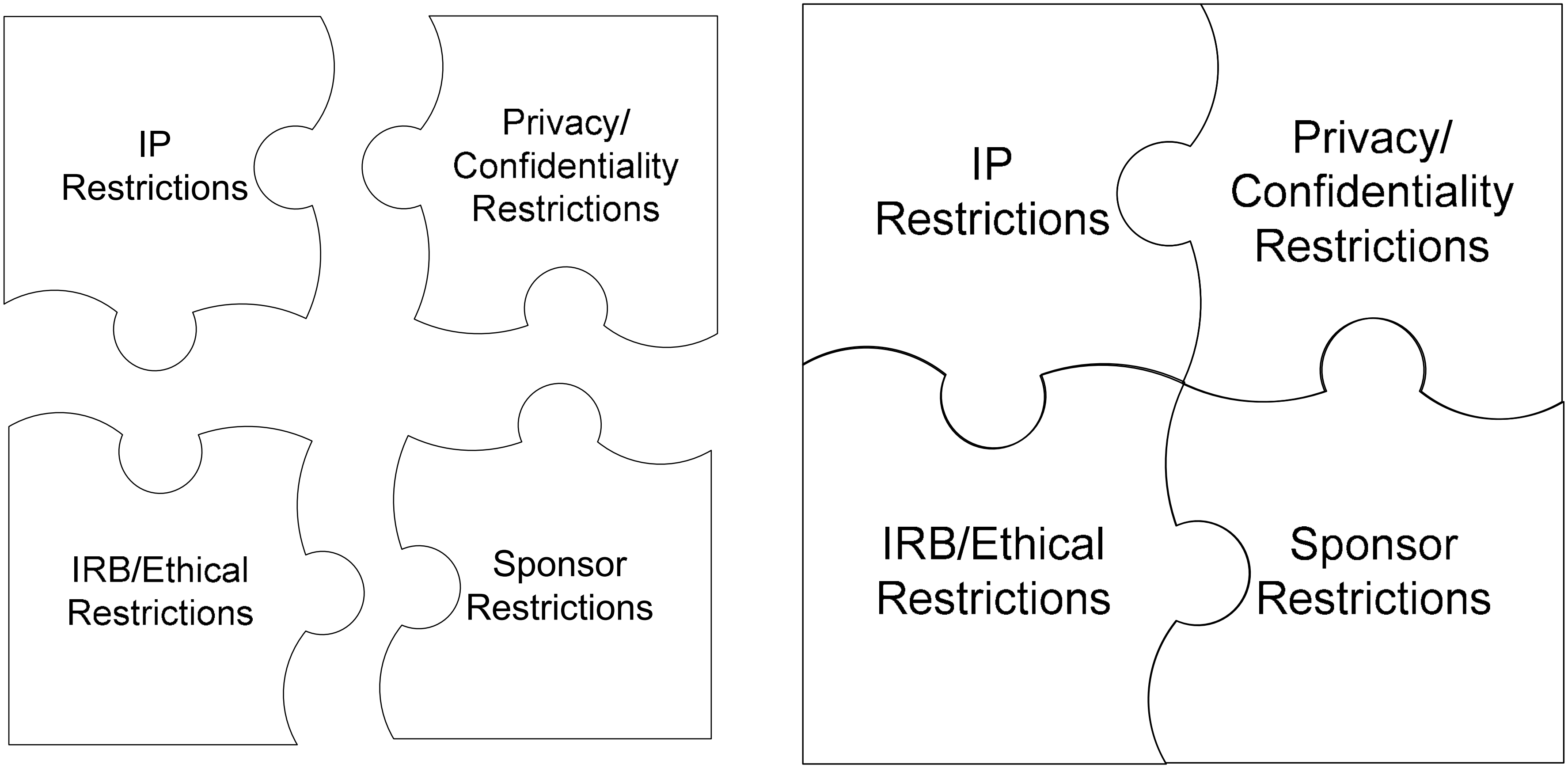


Low-Sensitivity Data

Moderate Sensitivity Data

High-Sensitivity Data

Pulling it all Together …



National Cancer Institute

caBIG™

cancer Biomedical Informatics Grid™

Other DSSF Resources

Other DSSF Resources

- Policies and Procedures
- Guidance and Best Practices
- Model Agreements and Other Documents
- Decision Support Tools

National Cancer Institute

Getting Connected with caBIG™

DATA SHARING AND SECURITY FRAMEWORK

Achieving technical interoperability in and of itself will not deliver the full benefits of accelerated translational research and expedited care for the patient. These are only possible when researchers and oncologists have the information they need to do their work better. However, there are very important non-technical considerations, such as patient privacy protections and intellectual property interests, which affect the ability to make data available.

The Data Sharing and Security Framework is designed to facilitate appropriate data sharing between and among organizations by addressing legal, regulatory, policy, proprietary, and contractual barriers to data exchange.

When fully developed, the Data Sharing and Security Framework will consist of policies, processes, model agreements, model data sharing plans and other materials that participating centers agree to help develop and to adopt as appropriate. Together, these documents will cover critical issues such as the creation of a trust fabric for accepting user authentication credentials from other institutions, standards for the levels of security that are needed for different types of data, and suggestions on how to share data subject to the Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security Rules, as well as other applicable laws, regulations, and policies.

This document provides an overview of the Data Sharing and Security Framework. It outlines what tools may be included in this bundle, their specific function, and the role participating cancer centers are expected to play.

The Data Sharing and Security Framework is part of the National Cancer Institute's overarching goal to connect the people, institutions, and data in the cancer community through caBIG™. This collection of tools and capabilities is one of three "bundles" that have been designed to help support and streamline clinical trials, imaging, tissue banking, and integrative cancer research, and to provide the materials needed to join the secure caBIG™ data-sharing framework.

Visit https://caBIG.nci.nih.gov/inventory for more detailed information and access to caBIG™ resources.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

caBIG

Guidelines for Preparing Data Sharing Plans

Purpose

- Provide a framework for organizing information about the data to be shared and the mechanism for sharing important to various data stewards or other interested institutional officials

Background information related to project in which caBIG™ infrastructure will be used

- Issues that drive legal/regulatory determinations
  - Summary of data elements, intended recipients, mechanisms for data sharing (access controls, agreements), timing, objectives of the project; and, who may have an interest in the data

Information about institutional units (including IRBs) that must approve data sharing plans

Open-ended questions regarding additional anticipated challenges

Model Informed Consent/Authorization

- **Model language re: caBIG™ for use in pre-existing authorization/consents**
  - Provide basic language for use by adopters and others to facilitate data sharing within their own documents; facilitate adoption of caBIG™ language in other models under development by various other groups

- **Standardized choices for research participants related to specimen use and/or data sharing**
  - Standardize choices in authorization/consent forms; facilitate adherence to patient/participant choices

- **Model informed consent and HIPAA authorization document – “the whole package” (disclosure + options)**
  - Assist institutions and smaller provider-based participants in drafting informed consent/authorization forms compliant with Common Rule, FDA and HIPAA requirements

- ***STATUS: Final drafts complete and will be posted for review in May 2008 for comments/feedback from the caBIG community, others.***

Guidelines for De-identifying Data

- **DSIC Workspace informational paper in development:**
  - rationale for de-identification
  - what de-identification encompasses
  - risks and benefits of sharing de-identified data
  - methods for de-identification
  - current technical approaches

- **Purpose: provide baseline information concerning de-identification approaches for institutions and entities responsible for overseeing human subjects research and protecting the privacy of the patient health data**

- **Next step -- develop practical guidelines on de-identification processes**

Ongoing DSIC Workspace Activities

- Inform requirements for caBIG™ tool development and adoption for compliance with caBIG principles
- Provide support to caBIG participants that develop, adopt, and utilize caBIG tools and infrastructure
- Prepare position statements and educational documents, including peer-reviewed publications, that describe views of caBIG community
- Develop responses to external policies and guidelines that may affect caBIG™ activities

National Cancer Institute

caBIG™

cancer Biomedical Informatics Grid™

Future State of DSSF

DSSF-Based Sensitivity Assessment

✓ DSSF Decision Support Tool

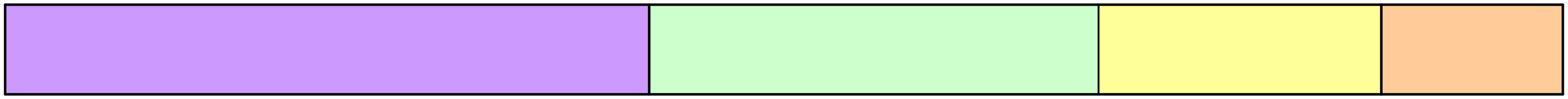
✓ Guidelines for Data Sharing Plans

✓ Model Informed Consent/HIPAA Authorization

□ De-identification Guidance

□ Citation Service/Tracker

□ Best Practices for Sharing Unpublished Data



Agreement Simplification

Medium sensitivity data – “yellow” lane

❑Standardized click-through agreement – medium sensitivity data

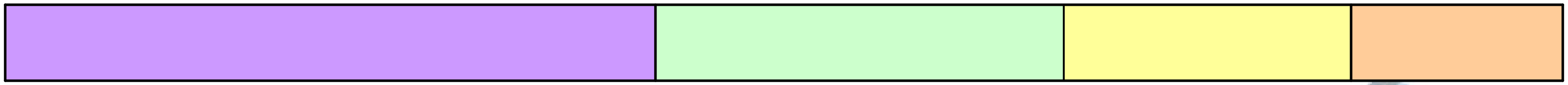
❑Technical implementation – medium sensitivity data

High sensitivity data – “orange” lane

❑Make terms of individual contracts accessible in response to data queries (one-to-many offers)-- initial technical implementation

❑Develop guidelines for developing data sharing agreements for high sensitivity data

❑Identify standardized contract terms for transactions involving high sensitivity data that can be "adopted" cafeteria-style in lieu of individually prepared contracts—second phase technical implementation



caBIG™ Trust Fabric

Non-sensitive data – “green” lane

- □ Authentication policies & procedures
  - ✓ Grid Host Agreement
  - ✓ Certificate Practice Statement
  - □ Grid User Agreement
  - □ Identity Provider Agreement

Low to medium sensitivity data – “yellow” lane

High sensitivity data – “orange” lane

- □ Authentication policies & procedures
  - □ Grid Host Agreement
  - □ Certificate Practice Statement
  - □ Grid User Agreement
  - □ Identity Provider Agreement
- □ Authorization policies/procedures

Best Practices for Sharing Unpublished Data

Resolve researcher and institutional concerns

- Engage key representatives from journals ranging the spectrum of biomedical research and to verify or repudiate assertions of researchers regarding policies about unpublished data.
- Engage representatives that participate in scientific review process/tenure review committees to discuss the impact of existing incentive structures around getting grants/tenure on data sharing
- Address issues of data provenance so that original collector continues to get attribution and funded institution can get metrics to justify value of devoting resources to data collection
- Generate report and best practices; disseminate at appropriate professional venues for scientists and university tech transfer/commercialization units.

The caBIG™ Initiative

Enterprise Support Network facilitates adoption

- **Knowledge Centers** – Provide domain-specific information and limited levels of support

- **Service Providers** – Offer technical and implementation support; software development; and documentation and training development and delivery

- **Program Offices** – Enable installation and operation of caBIG™ tools and infrastructure across multiple departments in individual institutions

caBIG™ Enterprise Support Network

National Cancer Institute

caBIG

caBIG Enterprise Support Network

NCI is Committed to Increasing caBIG™ Deployments

caBIG™ Funding for Next Generation Adoption

Learn More

How you can participate

- **Evaluate the use and integration of the DSSF into your Center's workflow; implement available DSSF tools; provide feedback to the caBIG Program on your experience with DSSF tools.**
  - Use the Framework to determine data set (and data element) sensitivity
  - Set levels of authentication required for general service and data access
  - Set levels of authorization required for individual data sets/data elements

- **Participate in caBIG™ Data Sharing and Intellectual Capital (DSIC) Workspace efforts (for Centers seeking greater input in developing and refining DSSF tools:**
  - Attend DSIC WS conference calls and F2F meetings
  - Review and comment on the proposed elements of DSSF Bundle (model documents, guidance/best practices and position papers:
    - ***How useful? What changes are needed for your institution?***
    - ***What issues do you face that DSIC can assist with?***

caBIG™: Getting Involved

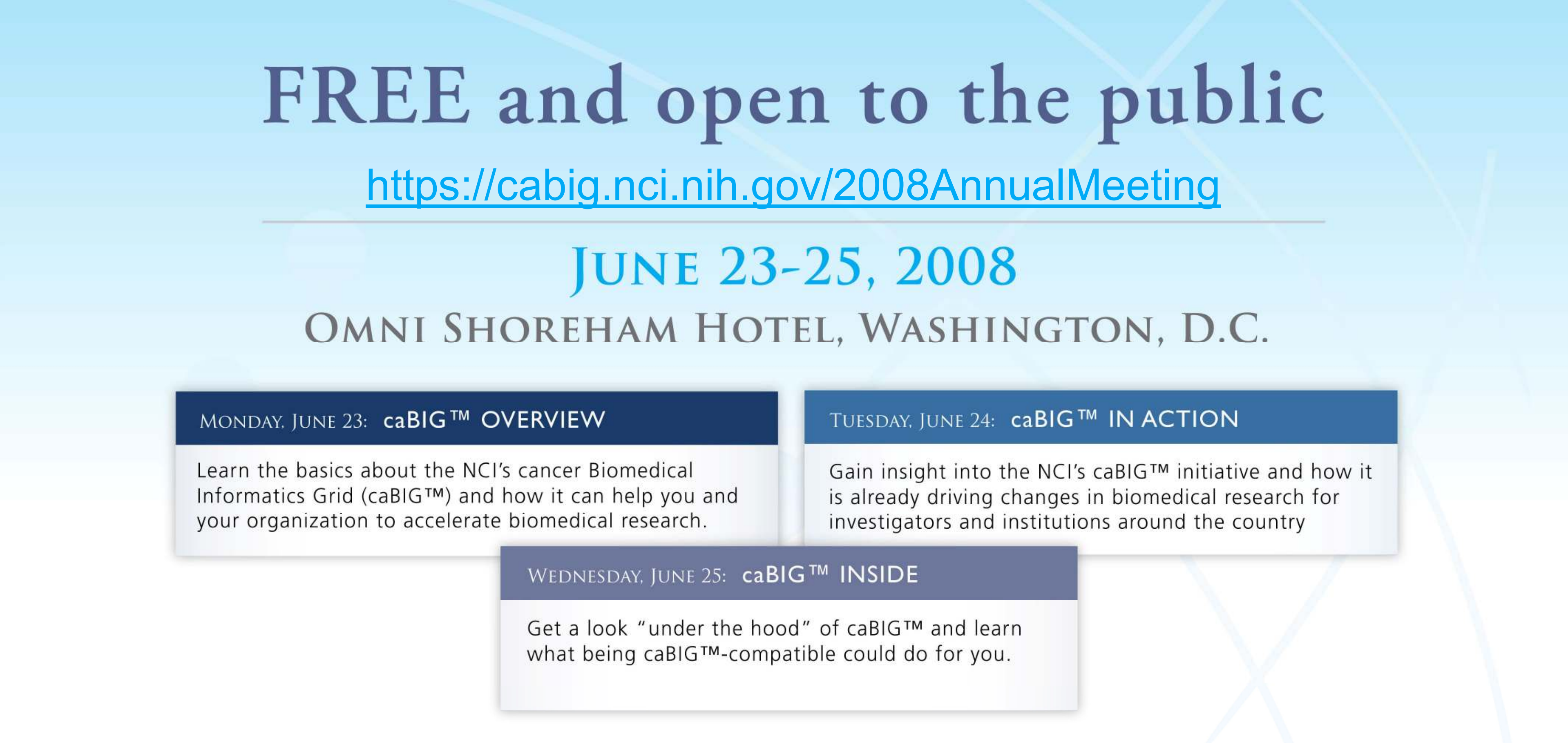
- Track caBIG™ activities on the caBIG™ website at https://cabig.nci.nih.gov/

- Attend the caBIG™ Annual Meeting in June 2008

- Sign up for the caBIG™ mailing list at http://list.nih.gov/archives/cabig_announce.html

- Join the DSIC Workspace at https://cabig.nci.nih.gov/working_groups/DSIC_SLWG/index_html

2008 caBIG™ Annual Meeting



FREE and open to the public

https://cabig.nci.ih.gov/2008AnnualMeeting

June 23-25, 2008

Omni Shoreham Hotel, Washington, D.C.

Monday, June 23: caBIG™ OVERVIEW

Learn the basics about the NCI's cancer Biomedical Informatics Grid (caBIG™) and how it can help you and your organization to accelerate biomedical research.

Tuesday, June 24: caBIG™ IN ACTION

Gain insight into the NCI's caBIG™ initiative and how it is already driving changes in biomedical research for investigators and institutions around the country

Wednesday, June 25: caBIG™ INSIDE

Get a look "under the hood" of caBIG™ and learn what being caBIG™-compatible could do for you.

Also join us for: caBIG™ World's Fair, Hack-a-thon, plenary speakers, demos, and more…

To subscribe to future updates, news, and case studies please visit: http://cabig.cancer.gov/email_signup.asp

Questions??

caBIG™: Power of Connection